This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representation of The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 780 777 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
25.06.1997 Bulletin 1997/26

(51) Int Cl.[6]: **G06F 17/30**, G11B 27/10

(21) Application number: **96309056.8**

(22) Date of filing: **12.12.1996**

(84) Designated Contracting States:
DE FR GB

(30) Priority: **21.12.1995 US 576106**

(71) Applicant: **Hewlett-Packard Company**
Palo Alto, California 94304 (US)

(72) Inventors:
- **Orsolini, Garry S.**
  Fair Oaks, California 95628 (US)
- **Bowers, Richard D.**
  Roseville, California 95747 (US)

(74) Representative: **Jehan, Robert et al**
Williams, Powell & Associates,
34 Tavistock Street
London WC2E 7PB (GB)

(54) Indexing of recordings

(57) A recording is indexed by keywords. In order to perform the indexing, an audio portion (12) of the recording is transcribed (31) to produce text in a text file. A time stamp (32) is associated with each word in the text. Each time stamp (32) indicates a time in the recording at which occurs an associated word. Once a recording has been indexed, the recording may be searched along with other recordings. For example, in response to a user choosing a keyword (46), a text file for each recording is searched for occurrences of the keyword (46). At the conclusion of the search, each recording which includes an occurrence of the keyword is listed (42). When a user selects (42) a first recording and a particular occurrence of the keyword (46), the first recording is played starting slightly before a time corresponding to a first time stamp associated with the particular occurrence of the keyword in the first recording. In response to control sequences, prior and next occurrences of the keyword (42) can be observed in one or multiple recordings.

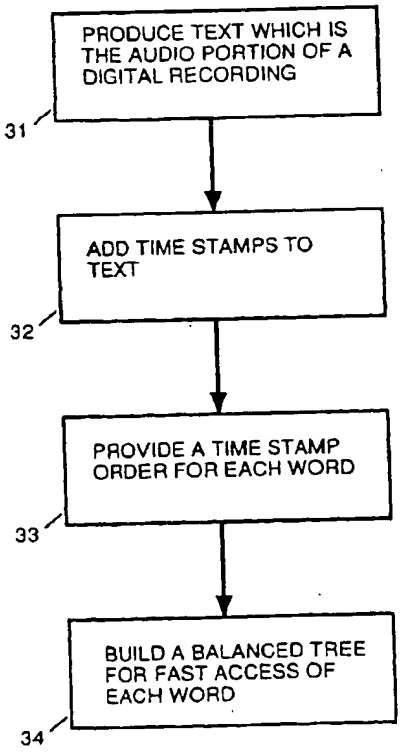


FIGURE 2

EP 0 780 777 A1

Printed by Jouve, 75001 PARIS (FR)

Description

The present invention relates to indexing of a recording for use, for example, in performing content searches within digital audio and audio-video recordings, using keywords to index digital audio and audio-video recordings.

Improvements in storage and compression technologies have allowed a revolution in multimedia. Audio recordings are now often stored in digital format. In addition, it is now feasible to convert full length movies into digital audio-video (video) recordings for replay. Using digital video, a user may, with off-the-shelf software products, access and edit full-screen, full-motion video recordings.

In order to make the best use of a computer's ability to manipulate digital audio and audio-video recordings, it is desirable to have some way to perform content searches. Currently, the ability to perform content searching is significantly limited or non-existent.

There exists some limited ability in the art to perform content searches of images. See for example, the QBIC Project by IBM Corporation, having a business address of 650 Harry Road, San Jose, California 95120. However, such searching of images on content is limited to visual content and is not capable of performing content searches on digital audio recordings.

The present invention seeks to provide indexing of recordings for performing content searches.

According to an aspect of the present invention there is provided a method of indexing a recording as specified in claim 1.

The method may index all text words.

Preferably, in response to a user choosing a keyword, the method searches the set of words for all occurrences of the keyword. Advantageously, this step includes listing all occurrences of the found keyword and, in response to selection of an occurrence of the listed keyword, playing the recording, preferably starting slightly before a time corresponding to a time stamp associated with the selected occurrence of the keyword.

The method is applicable to various types of recording, including audio and audio-video recordings.

According to another aspect of the present invention there is provided a method of accessing selections within a plurality of recordings as specified in claim 7.

According to another aspect of the present invention there is provided a system for accessing selections within a plurality of recordings as specified in claim 10.

It will also be apparent that the invention extends to a system for indexing a recording.

In the preferred embodiment, a recording is indexed by keywords. In order to perform the indexing, an audio portion of the recording is transcribed to produce text in a text file. The transcription may be performed, for example, manually by a transcriber or using speech recognition technology. After transcription, a time stamp is associated with each word in the text. Each time stamp indicates a time in the recording at which occurs an associated word. The time stamps may be added to the text file, for example using speech recognition technology. Alternately, the time stamps may be added to the text file by an operator using a computing system. For example, the text is displayed in a first window of a computer display. The recording is displayed in a second window of the computer display. Upon the operator selecting (or upon automatic selection of) a selected word of the text in the first window, a time stamp is added to the text file which indicates an elapsed time from a beginning of the recording until selection by the operator of the selected word. Once the operator has in this way or by some other method assigned time stamps to a subset of words in the text, interpolation may be used to assign time stamps to the remaining words in the text which are not within the subset of words assigned time stamps by the operator. Once time stamps have been assigned to each word in the text, the words and associated time stamps may be arranged in a balanced tree for efficient access by a search program. Other search techniques can be used instead of the balance tree. For example, a binary tree can be used.

The preferred embodiment also provides for keyword searching of a plurality of recordings, each with an associated text file created as described above. In response to a user choosing a keyword, a text file/balanced tree for each recording is searched for occurrences of the keyword. At the conclusion of the search, each recording which includes an occurrence of the keyword is listed. When a user selects a first recording and a particular occurrence of the keyword, the first recording is played starting slightly before a time corresponding to a first time stamp associated with the particular occurrence of the keyword in the first recording.

For example, after searching on a keyword, the recordings may be listed as follows. The list of recordings which include an occurrence of the keyword are displayed in a first window of a computer display. One of the recordings from the list of recordings displayed in the first window is highlighted. A user may select which recording is highlighted. In one embodiment, upon a user selecting a particular recording, a first-in-time occurrence of the keyword within the particular recording is played. Keystroke commands may be used to jump to other occurrences. In an alternate embodiment, when a recording displayed in the first window is highlighted, each of the occurrences of the keyword within the highlighted recording is listed. This may be done, for example, in a second window of the computer display.

Various variations may be made to the preferred embodiments. For example, in addition to allowing searching on a single keyword, searching may be performed on multiple keywords connected by Boolean logic or may be performed on concepts. It is also envisaged that the process may be entirely automated in some applications.

The present invention can allow for efficient content

searching of recordings, improving over other currently available schemes to index recordings.

An embodiment of the present invention is described below, by way of example only, with reference to the accompanying drawings, in which:

Figure 1 illustrates steps taken to allow keyword indexing of digital recordings in accordance with the preferred embodiment.

Figure 2 is a flowchart which shows steps by which text for a digital recording is keyword indexed in accordance with the preferred embodiment.

Figure 3 and Figure 4 show computing displays which illustrate the preparation of a data base used for keyword indexing of digital recordings in accordance with the preferred embodiment.

Figure 5 shows a computing display used for keyword index searches of a video library in accordance with the preferred embodiment.

Figure 6 shows a computing display used for keyword index searches of a video library in accordance with an alternate embodiment.

Figure 1 illustrates steps taken to allow keyword indexing of digital recordings. A recording source 11 is digitized and compressed to produce digitized recording file 13. Recording source 11 is, for example, an audio recording or an audio-video recording. When recording source 11 is an audio-video recording, data in digitized recording file 13 is, for example, stored in MPEG-1 format. Digitized recording file 13 may be produced from analog recording source 11 using, for example, OptiVideo MPEG 1 Encoder available from OptiVision, having a business address of 3450 Hillview Ave., Palo Alto, CA 94304.

In addition, the audio portion of recording source 11 is transcribed to produce a text file 12 which includes the text. The transcription may be performed manually. Alternately, the audio portion of recording source 11 may be transcribed directly from recording source 11 or digitized recording file 13 using computerized speech recognition technology such as DragonDictate for Windows available from Dragon Systems, Inc., having a business address of 320 Nevada Street, Newton, MA 02160. Text file 12 and digitized recording file 13 are then made available to a computer system 14.

Figure 2 is a flowchart which shows steps by which text for a digitized recording file 13 is keyword indexed. In a step 31, text is produced which is the audio portion of digitized recording file 13. This text is a result of the transcription described above.

Figure 3 illustrates the result of the transcription process. Figure 3 shows a window 23 in a computer screen 21. Within window 23 is the transcribed text of the audio portion of recording file 13.

In a step 32, shown in Figure 2, time stamps associated with words in the text are added to the transcribed text. In the preferred embodiment, the time stamps are in milliseconds and indicate elapse of time relative to the starting point of the digital recording within recording file 13.

Placement of time stamps may be performed, for example, with the help of an operator utilizing, on computer 14 (shown in Figure 1), software specifically designed to add time stamps. For example, the recording is played by computer 14. For an audio-video recording, a window 22 in computer screen 21, as shown in Figure 3, may be added in which the audio-video recording is played. The operator of computer 14, using cursor 24, selects words as they are spoken in the recording played by computer 14. Whenever the operator selects with cursor 24 a word from the text in window 23, the software running on computer 14 time stamps the word with the current time duration which represents the elapse of time relative to the starting point of the digital recording.

Figure 4 further illustrates this process. In Figure 4, time stamps TS1, TS2 and TS3 have been added to text 23 by an operator as described above. Source code for software which implements the time stamp feature discussed above for audio-video recordings will be apparent to the skilled person. Alternately, step 32, shown in Figure 2, may be automated so that speech recognition technology is used to trigger the placement of time stamps within text 23.

After the time stamps have been added to text 23, in a step 33 shown in Figure 2, every word of text 23 is assigned a time stamp. For words which were not assigned a time stamp in step 32, interpolation is used to determine an appropriate time stamp.

For example, Table 1 below shows a portion of text 23 after the completion of step 32.

Table 1

Once::11 upon a time: :20 there was a boy::28 named Fred. He went::35 to the forest::44. . . .

In the example given in Table 1, the word "Once" was spoken at 11 milliseconds from the beginning of the audio track of the digital recording. The word "time" was spoken at 20 milliseconds from the beginning of the audio track of the digital recording. The word "boy" was spoken at 28 milliseconds from the beginning of the audio track of the digital recording. The word "went" was spoken at 35 milliseconds from the beginning of the audio track of the digital recording. The word "forest" was spoken at 44 milliseconds from the beginning of the audio track of the digital recording.

In order to assign time stamps to the remainder of the words, interpolation is used. For example, nine milliseconds elapsed between the word "Once" and the word "time". There are two words, "upon", and "a", which occur between "Once" and "time". As a result of the interpolation, the words "upon", and "a" are assigned time stamps of 14 milliseconds and 17 milliseconds, respectively. This is done so that there is allocated three milliseconds between the occurrence of the word "Once" and the word "upon"; there is allocated three millisec-

onds between the occurrence of the word "upon" and the word "a"; and there is allocated three milliseconds between the occurrence of the word "a" and the word "time".

The words and their time stamps are placed in an output file. For example, the output file may have on each line a single word, separated by a tab character from a time stamp for the word. Table 2 below shows the form of the file for the example text file shown in Table 1 above:

Table 2

Once	11
upon	14
a	17
time	20
there	22
was	24
a	26
boy	28
named	30
Fred	32
He	34
went	35
to	38
the	41
forest	44
.	
.	
.	

Source code for software which implements the interpolation feature discussed above will be apparent to the skilled person. Alternately, in step 32 every word may be assigned a time stamp, for example using speech recognition technology, so that no interpolation is necessary. Using speech recognition technology, words may be transcribed and time stamped simultaneously. Alternately, speech recognition technology may be used in a separate pass in which time stamps are added to a transcription of the text. When used in a separate pass to add time stamps to words in a text, the speech recognition software adds time stamps for unrecognized words by interpolation, as described above. In a step 34 shown in Figure 2, a balanced tree is built which allows fast access of words within the output file. The balanced tree is built, for example, using an algorithm known in the art. See for example, Robert Sedgewick, "Algorithms in C++", Addison-Wesley Publishing Company, 1992, pp. 215-229. Appendix C includes source code for software which implements the construction of the balanced tree as set out in step 34. Alternately, a binary tree or other searching algorithm may be used. In other embodiments, searching may be performed directly on the output file constructed in step 33.

The balanced tree constructed in step 34 serves as a keyword index of the digital recording. The balanced tree is accessed to locate where a word is spoken in the movie.

For example, Figure 5 illustrates an interface on a computer screen 51 which utilizes the keyword index constructed as described above. In a box 56, a user types one or more keywords connected by Boolean variables. In a window 52, recordings are listed in which the keyword(s) appear. The number of "hits" of a keyword appears is listed next to the recording. In the preferred embodiment, the recordings are listed in descending order by the number of keyword occurrences. A user selects a recording using cursor 54, cursor keys, or some other way. When a recording is selected, for example using an "OK" button 58 by the user, the portion of the selected recording (listed in window 52) in which the first occurrence of the selected keyword appears is played. For an audio-video recording, the visual portion is displayed in display window 55. The portion of the recording is displayed for a configurable duration (e.g., two seconds) starting, for example, one second before the occurrence of the keyword. Using keyboard commands, a user can continue viewing the recording, fast forward, reverse, skip to the next occurrence of the keyword, go back to the last occurrence of the keyword, continue playing and so on. The interface also includes a "cancel" button 59. Source code for software which (in addition to implementing the construction of the balanced tree) implements the keyword searching of recordings, as discussed above will be apparent to the skilled person.

The interface in Figure 5, may be enhanced to include additional features. For example, Figure 6 illustrates an interface on a computer screen 41 which utilizes the keyword index constructed as described above. In a box 46, a user types one or more keywords connected by Boolean variables. In a window 42, recordings are listed in which the keyword(s) appear. The number of times a keyword appears in a recording is listed next to the recording. In the preferred embodiment, the recordings are listed in descending order by the number of keyword occurrences. A user selects a recording using cursor 44, cursor keys, or some other way. When a recording is selected, in a window 43, a list of all the occurrences for the keyword(s) are listed. In one embodiment, a fragment of text, along with the time stamp, are displayed for each occurrence. Alternately, only the keyword and time stamp, or only the time stamp are displayed for each occurrence.

Upon selection of an "OK" button 48 by the user, the portion of the selected recording (in window 44) in which the selected keyword (in window 43) appears is played in a digitized recording display window 45. The portion of the recording is displayed for a configurable duration (e.g., two seconds) starting, for example, one second before the occurrence of the keyword. Using a control panel 50, a user can continue viewing the recording, fast forward, reverse, skip to the next occur-

rence of the keyword, go back to the last occurrence of the keyword, continue playing and so on. The interface also includes a "cancel" button 49.

In addition to searching on one or more keywords connected by Boolean variables, the balanced tree formed in step 34 (shown in Figure 2) may also be searched using concept based searching techniques, for example using Metamorph available from Thunderstone Software-EPI, Inc. having a business address of 11115 Edgewater Drive, Cleveland, Ohio 44102.

The foregoing discussion discloses and describes merely exemplary methods and embodiments.

The disclosures in United States patent application no. 08/576,106, from which this application claims priority, and in the abstract accompanying this application are incorporated herein by reference. The US parent application also includes examples of the source codes mentioned herein.

Claims

1. A method of indexing a recording comprising the steps of:

   (a) transcribing (31) an audio portion of the recording to produce text in a text file; and,
   (b) providing (32) for each of a set of words in the text, a time stamp which indicates a time in the recording at which each word in the set of words occurs.

2. A method as in claim 1 wherein step (a) is accomplished manually by a transcriber or with the use of speech recognition technology.

3. A method as in claim 2 wherein when step (a) is accomplished with the use of speech recognition technology, steps (a) and (b) are performed simultaneously.

4. A method as in claim 1, 2 or 3, wherein step (b) includes the substeps of:

   (b.1) providing for each of a subset of the set of words in the text, a time stamp which indicates a time in the recording at which each word in the subset of the set of words occurs; and,
   (b.2) for a remainder of the set of words which are not in the subset of the set of words, using interpolation to provide a time stamp which indicates a time in the recording at which each word in the remainder of the set of words occurs.

5. A method as in claim 4 wherein the recording is an audio-video recording and wherein substep (b.1) includes the substeps of:

   (b.1.1) displaying the text in a first window (23) of a computer display;
   (b.1.2) playing a video portion of the recording in a second window (22) of the computer display; and,
   (b.1.3) upon an operator selecting a selected word of the text in the first window, adding a time stamp (TS1...) to the text file which indicates an elapsed time from a beginning of the recording until selection by the operator of the selected word.

6. A method as in any preceding claim, comprising the step of:

   (c) arranging the set of words and associated time stamps into a balanced tree based on occurrences of each word in the set of words.

7. A method of accessing selections within a plurality of recordings, comprising the steps of:

   (a) in response to a user choosing a keyword, searching a plurality of text files for occurrences of the keyword, wherein text files are associated with recordings so that for each of the plurality of recordings, one text file from the plurality of text files includes a text of an audio portion of the recording, each word in each text file being associated with a time stamp (TS1...) which indicates an approximate location in an associated recording of an occurrence of the word;
   (b) listing (44) recordings which include an occurrence of the keyword; and,
   (c) upon a user selecting a first recording and a particular occurrence of the keyword, playing the first recording starting slightly before a time corresponding to a first time stamp associated with the particular occurrence of the keyword in the first recording.

8. A method as in claim 7 wherein in step (c) upon a user selecting the first recording, a first-in-time occurrence of the keyword within the first recording is automatically selected as the particular occurrence of the keyword.

9. A method as in claim 7 or 8 wherein step (b) includes the substeps of:

   (b.1) listing in a first window the recordings which include an occurrence of the keyword;
   (b.2) highlighting one of the recordings from the recordings listed in the first window; and,
   (b.3) listing each of the occurrences of the keyword within the recording highlighted in substep (b.2).

10. A system for accessing selections within a plurality

of recordings, comprising:

a plurality of text files (24), each text file including a text of an audio portion of an associated recording from the plurality of recordings;
search means for searching the text files for occurrences of the keyword in response to selection of a keyword; and
recording play means for playing the first recording starting slightly before a time corresponding to the particular occurrence of the keyword in the first recording in response to selection of a particular occurrence of the keyword within a first recording.

11. A system as in claim 10 wherein the search means includes a first keyword display (46) able to accept from a user a specification of a particular keyword; and a first window (42) operable to display a list of recordings which include an occurrence of the particular keyword.

12. A system as in claim 11 wherein the search means includes a second display (42) operable to display occurrences of the keyword within a recording highlighted in the first window.

13. A system as in claim 10, 11 or 12, wherein the search means is operable to search on a plurality of keywords connected by Boolean logic and/or to perform concept based searches on the keyword.

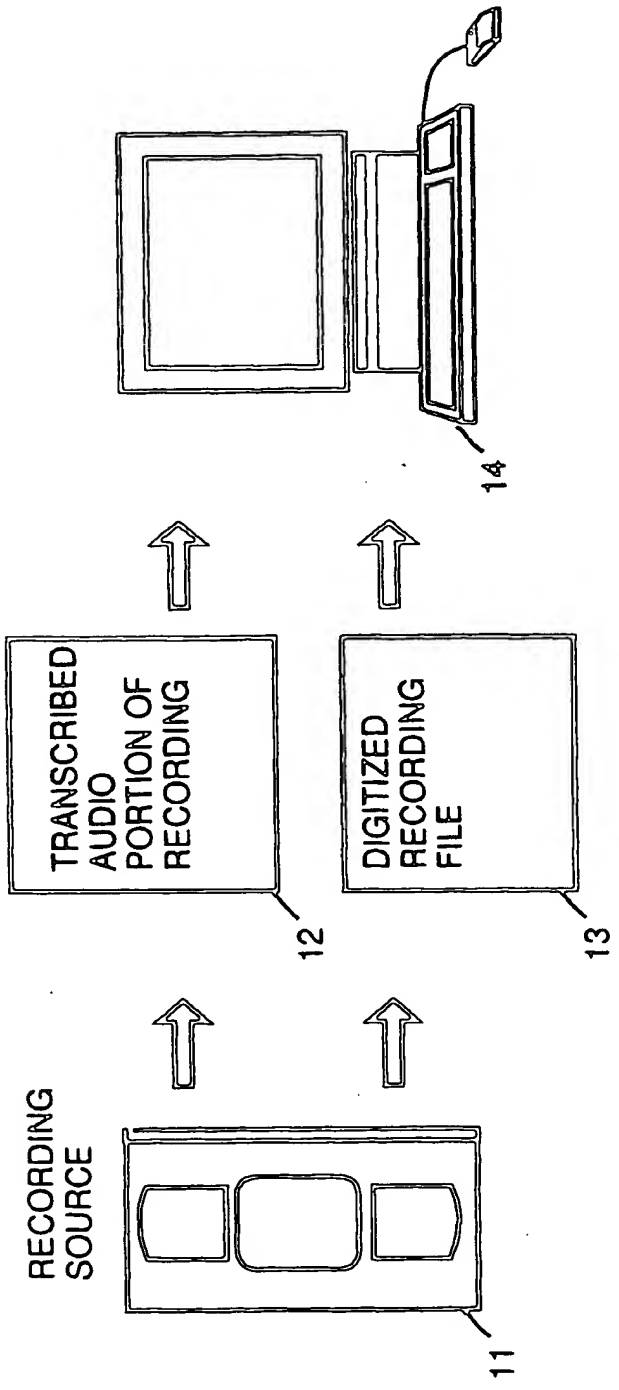

RECORDING SOURCE
11
TRANSCRIBED AUDIO PORTION OF RECORDING
12
DIGITIZED RECORDING FILE
13
14

FIGURE 1

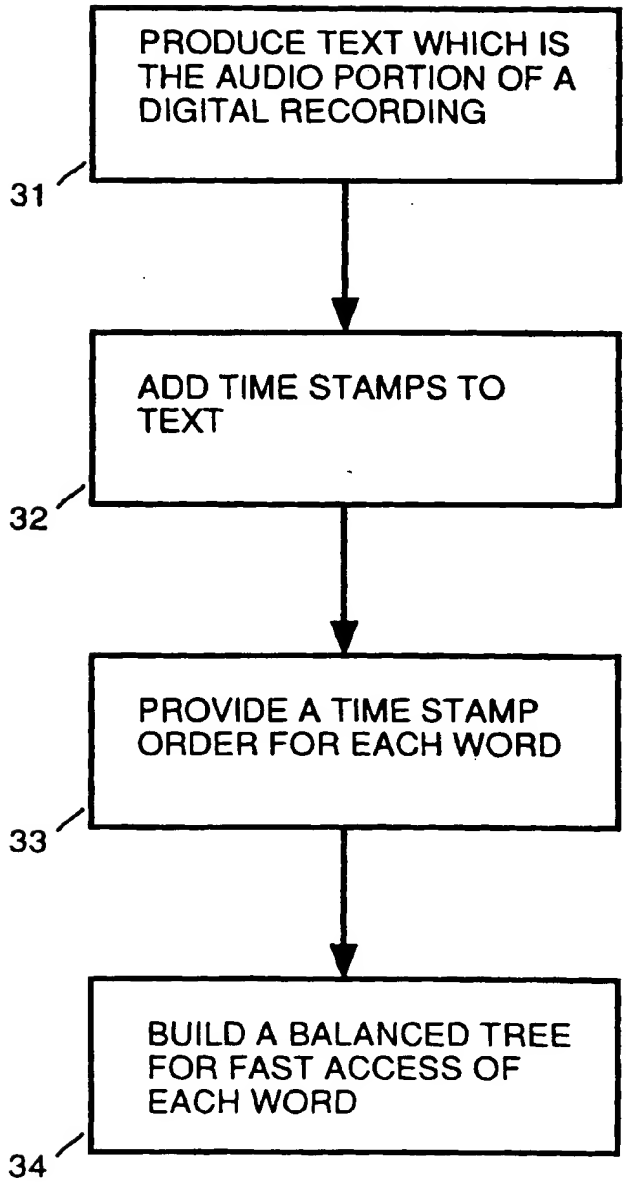

PRODUCE TEXT WHICH IS THE AUDIO PORTION OF A DIGITAL RECORDING
31
ADD TIME STAMPS TO TEXT
32
PROVIDE A TIME STAMP ORDER FOR EACH WORD
33
BUILD A BALANCED TREE FOR FAST ACCESS OF EACH WORD
34

FIGURE 2

FIGURE 3

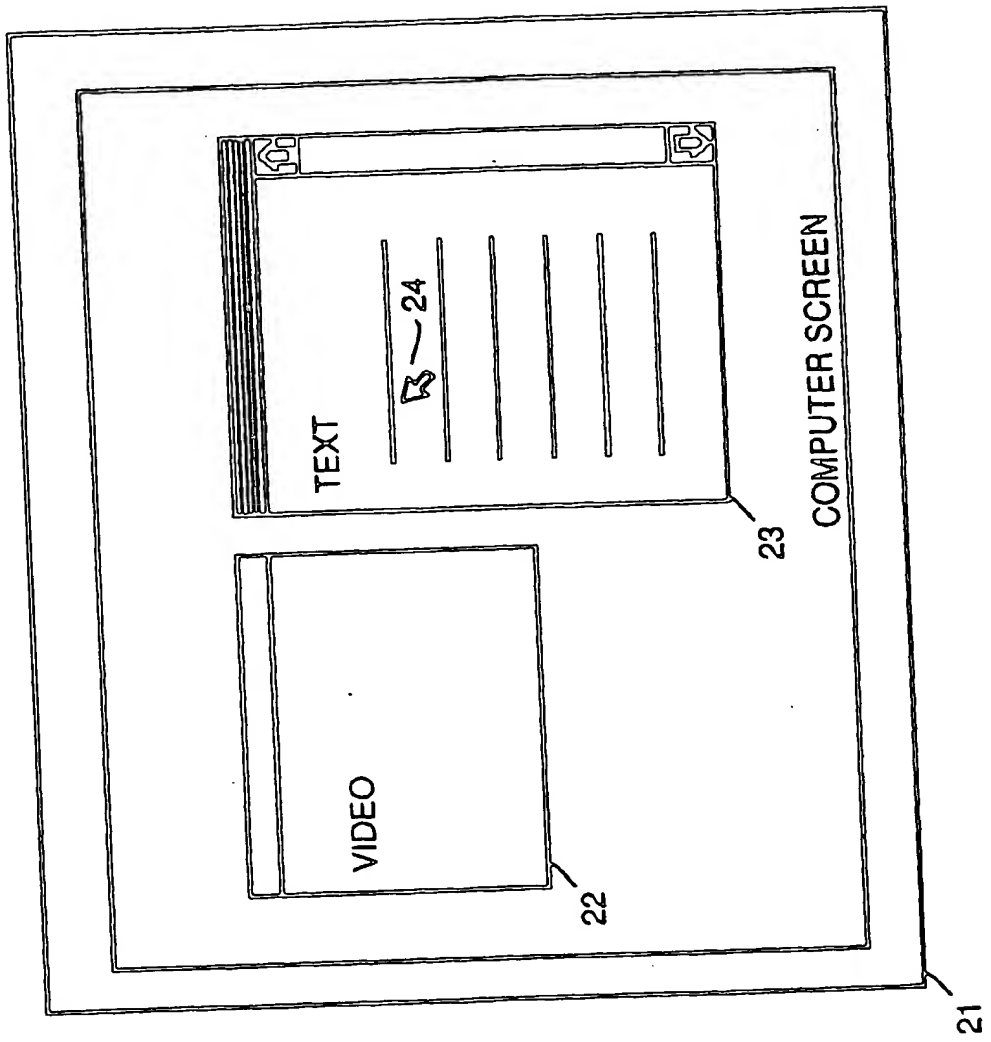
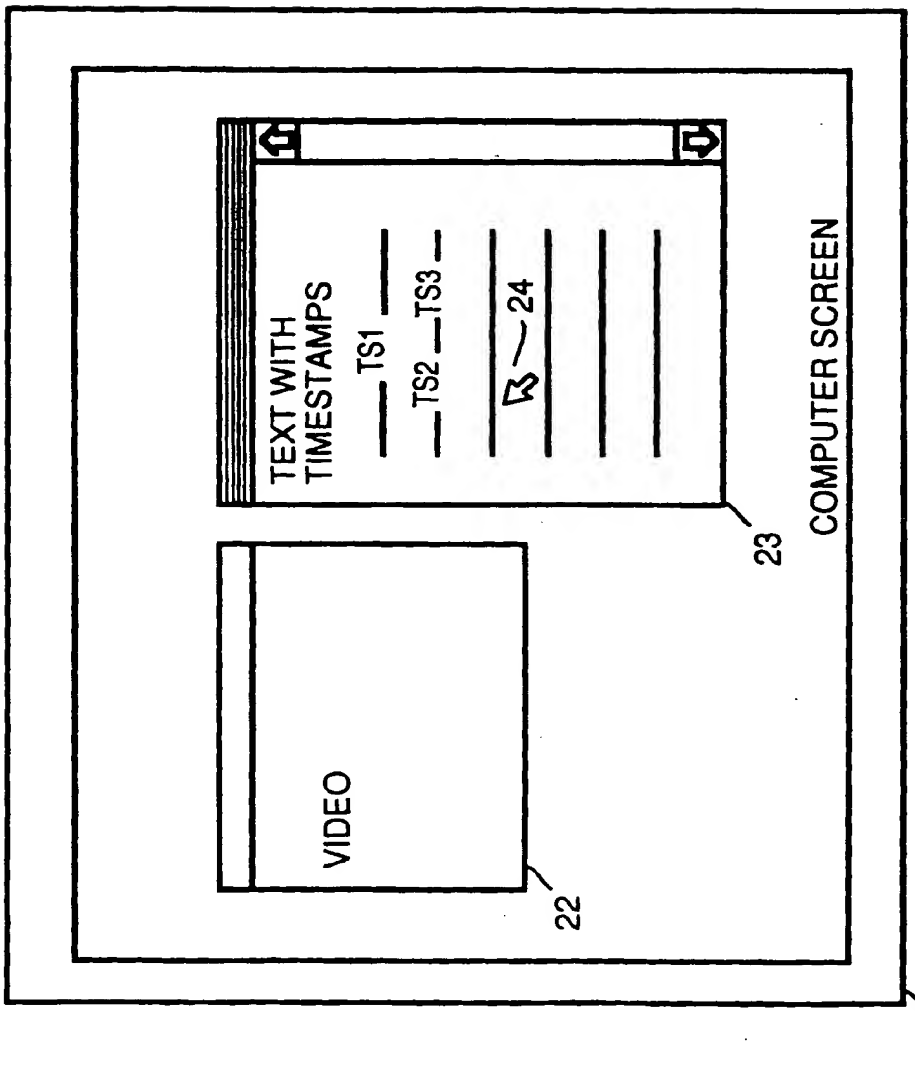


FIGURE 4


COMPUTER SCREEN
TEXT WITH TIMESTAMPS
TS1
TS2
TS3
24
23
VIDEO
22
21

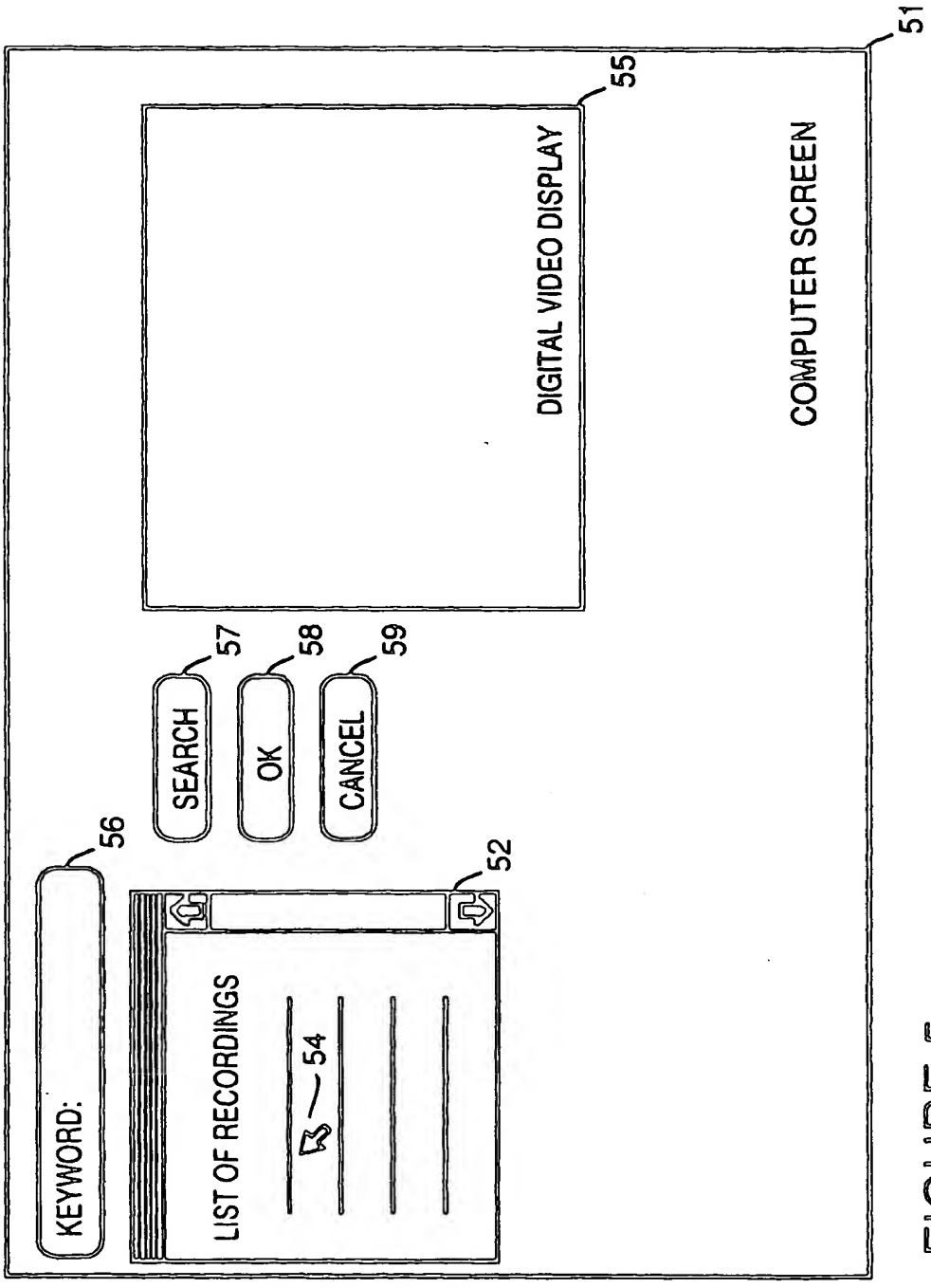

51
COMPUTER SCREEN
55
DIGITAL VIDEO DISPLAY
57
SEARCH
58
OK
59
CANCEL
56
KEYWORD:
52
LIST OF RECORDINGS
54

FIGURE 5

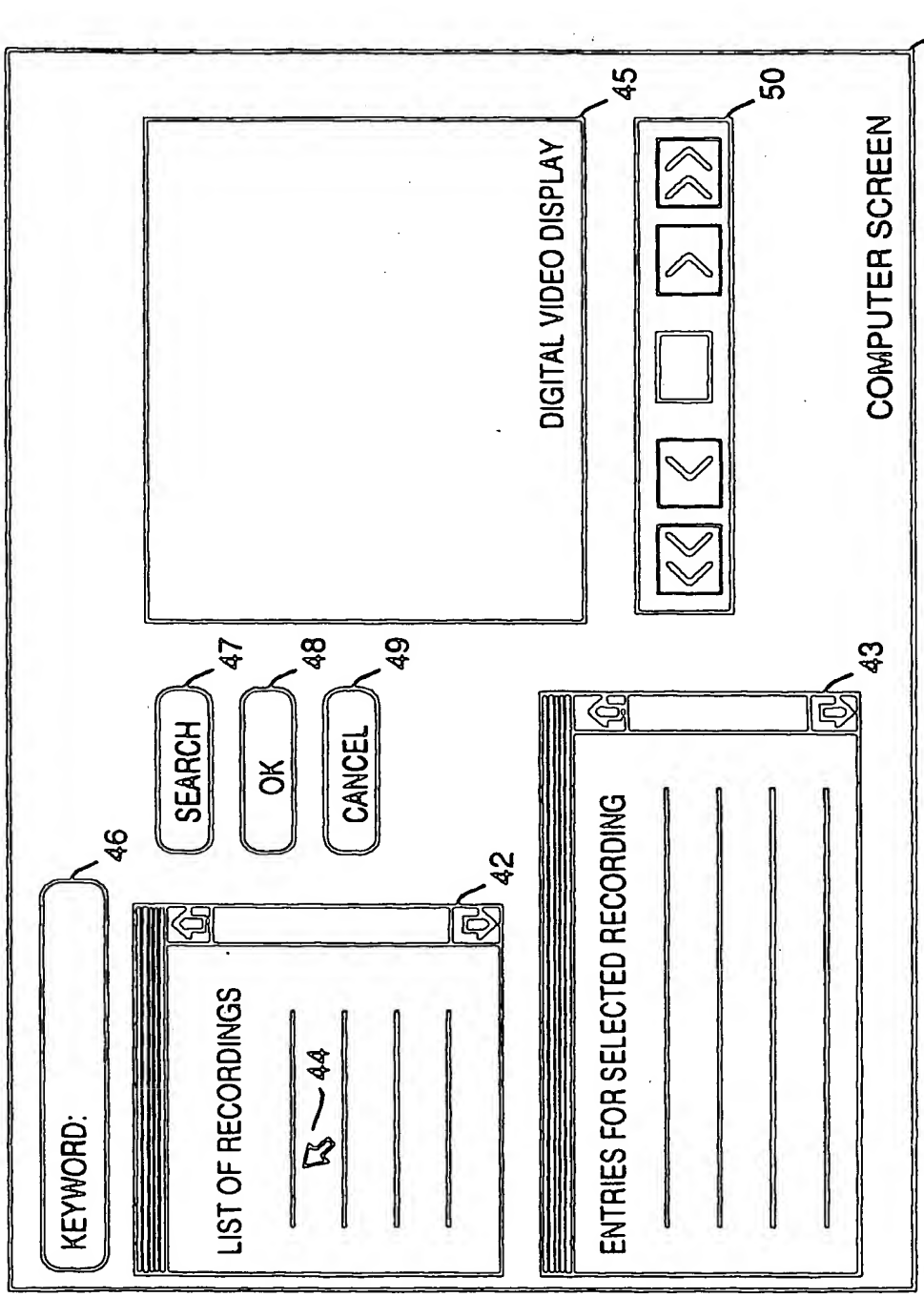

41
COMPUTER SCREEN
45
DIGITAL VIDEO DISPLAY
50
43
ENTRIES FOR SELECTED RECORDING
42
LIST OF RECORDINGS
44
46
KEYWORD:
47
SEARCH
48
OK
49
CANCEL

FIGURE 6

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 30 9056

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	EP 0 507 743 A (STENOGRAPH CORP) 7 October 1992 * page 2, line 23 - line 48 * * page 5, line 25 - line 53; figures 5,6 * * page 6, line 14 - line 36 *	1-4,7-13	G06F17/30 G11B27/10
X Y A	EP 0 649 144 A (IBM) 19 April 1995 * column 4, line 5 - column 6, line 14; figures 1,2 * * column 13, line 37 - column 14, line 8; figure 5 * * column 6, line 18 - line 46; figure 3 *	1-5 6 7,10	
X	US 5 136 655 A (BRONSON BARRY S) 4 August 1992 * column 2, line 67 - column 3, line 42; figure 1 *	1-3	
Y	PATENT ABSTRACTS OF JAPAN vol. 013, no. 590 (P-984), 26 December 1989 & JP 01 253027 A (NIPPON TELEGR & TELEPH CORP), 9 October 1989, * abstract *	6	TECHNICAL FIELDS SEARCHED (Int.Cl.6) G06F G11B
X	WO 92 11634 A (ARDIS PATRICK M ;MARKOVICH MARKO R (US); THOMPSON KEVIN W (US)) 9 July 1992 * abstract; figure 3 * * page 9, line 21 - page 11, line 22 *	1-3,10	
X	IBM TECHNICAL DISCLOSURE BULLETIN, vol. 33, no. 10A, 1 March 1991, page 295/296 XP000110048 "CORRELATING AUDIO AND MOVING-IMAGE TRACKS" * the whole document *	1,2,10	

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
BERLIN	1 April 1997	Deane, E

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 30 9056

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
P,X	COMPUTER, vol. 29, no. 5, 1 May 1996, pages 46-52, XP000591712 WACTLAR H D ET AL: "INTELLIGENT ACCESS TO DIGITAL VIDEO: INFORMEDIA PROJECT" * page 47, left-hand column, line 12 - right-hand column, line 36 *	1-3	
P,X	WO 96 12239 A (UNIV CARNEGIE MELLON) 25 April 1996 * page 7, line 17 - page 9, line 22 * * page 13, line 15 - page 14, line 27; figure 2 *	1-3	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
BERLIN	1 April 1997	Deane, E

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)